

# Rethinking Enterprise Search in a Web 2.0 World

#### Jayne Dutra

Enterprise Information Architect
DAMA International Symposium and Wilshire Metadata
Conference

Jet Propulsion Laboratory
California Institute of Technology
NASA
March 19, 2008









## Agenda

- Web 2.0 Meanings and Implications
- Search 1.0 So Yesterday!
- New Technologies
- Search 2.0
- Re-Imagining Enterprise Search
- Role of Data Reconciliation
- Next Generation and New Ideas







## New Developments: Web 2.0

#### Web 3.0

ca. 2005-2020

Web 2.0

ca. 2000-2010

## Web 1.0 ca. 1990-2000

• Portals

- \_ .
- Web Search Engines
- Web Sites
- Databases/ File Servers

Mash Ups

- Semantic Search
- Second Life, Avatars
- Thesauri,
   Taxonomies
- Blogs, Wikis, RSS
- Mobile Connections
- Social Networks, Social Bookmarking
- Instant Messaging
- Community Portals, Visualization

With thanks to Nova Spivak, Radar Networks and Mills Davis, Project 10X







## New Developments: Web 2.0

### Web 3.0

ca. 2005-2026

Seamless Integration

Projected Personae

 Ubiquitous Virtual Presence

#### Web 2.0

ca. 2000-2010

- Publishing to the Web
  - Available to the Many
  - Increased Connected-ness
- The Rise of the Wisdom of Crowds

## Web 1.0 ca. 1990-2000

- Publishing to the Web
  - -Controlled by a Few
  - -Complicated, expensive technologies

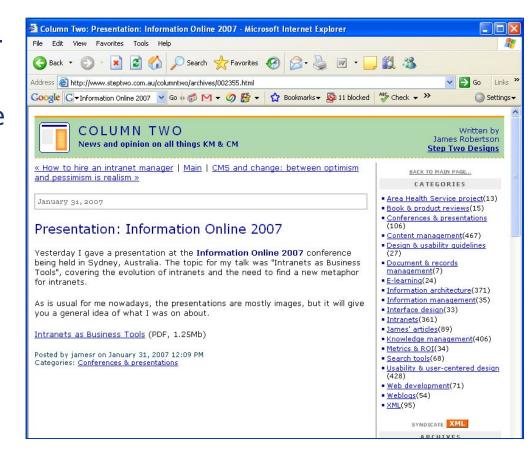






#### The Socialization of the Web

- Increased connectedness of individuals
  - Blogs and wikis enable increased content and knowledge sharing
  - Enriching user experience
  - Example: James
     Robertson blog on
     Information Online
     Conference 2007 in
     Sydney, Australia

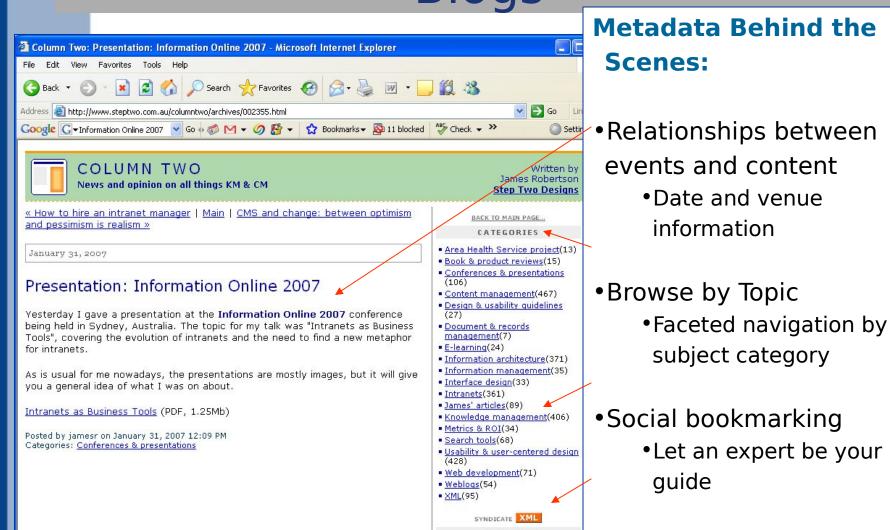








# The Socialization of the Web: Blogs









### Wisdom of Crowds

New publishing methods that utilize shared spaces

 Wikis, Facebook, MySpace, YouTube, Flickr, del.icio.us, Twitter, Wikipedia, etc., etc.

Need we say more?

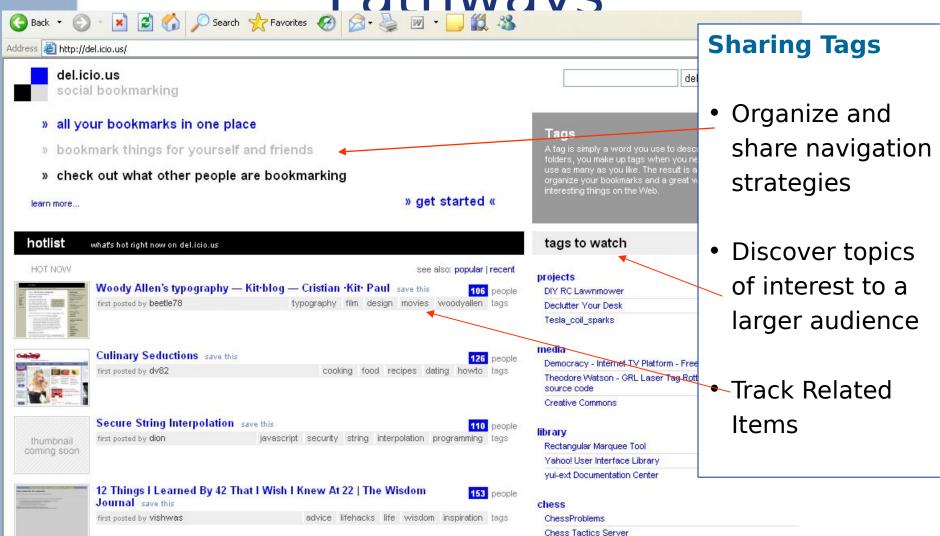
This has produced an explosion of content!







# Social Navigation Pathways









# Tagging as a Social Activity

Search Search Favorites 🚱 🔯 🔻 🐷 🕶 🗸 🔉	
io.us/tag/	Go Lin
del.icio.us / tag /	popular   recent login   register   help
Popular tags on del.icio.us	del.icio.us 🗸 search

This is a tag cloud - a list of tags where size reflects popularity. sort: alphabetically | by size

net 2008 advertising ajax apple architecture **art** article au audio bandom bandslash **blog** blogging **blogs** books business community computer cooking cool **CSS** culture **design** development diy download downloads **education** electronics entertainment environment fanfic fic film finance firefox flash food forum **free** freeware fun funny gallery game **games google** graphics gtd hardware health history home **howto** humor illustration images imported **inspiration** interesting internet java javascript jobs knitting language learning library lifehacks **linux** literature **mac** marketing math mor media microsoft mobile movies mp3 **music** network **news** online opensource osx photo **photography** photos photoshop php **politics** portfolio productivity **programming** python radio rails recipe recipes **reference** religion research resources ruby school science search security sga **shopping** slash social **software** statistics tech technology tips **tools** toread **travel tutorial** tutorials to ubuntu usa **video** visualization **web web2.0 webdesign** webdev wedding wiki windows wordpress writing yahoo youtube





#### Why is Like.com Different?

Like is a visual shopping engine that lets you find items by color, shape and pattern.

Click on Likeness Search to get started

#### Visual Search: Like.com

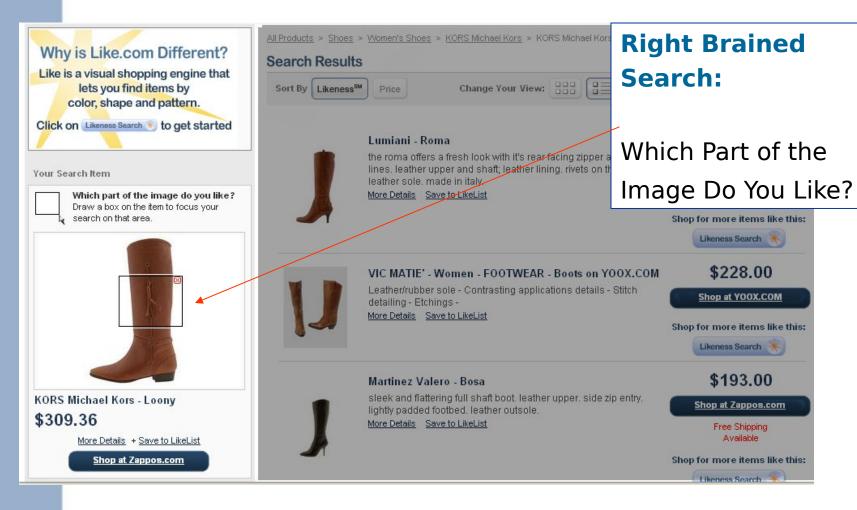








# Intuitive Visual Cues Instead of Keywords

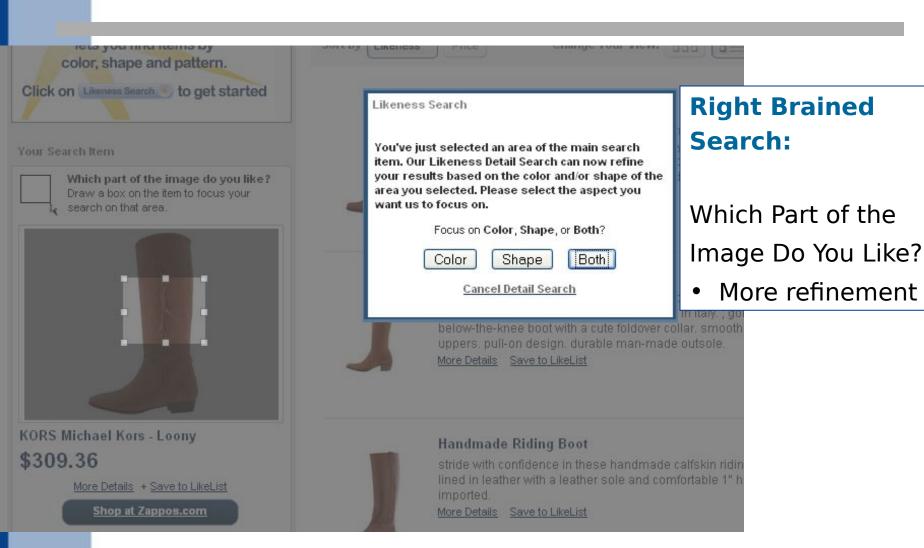








### Navigation by Image Selection









#### Search Evolution

Okay, that was on the Internet,
But what about **inside** the Enterprise?
Yep, we've got it.

Blogs, wikis, community portals, folksonomies, the need for increased personal dashboard reporting, RSS subscription, shared searches across repositories, all kinds of data formats and content items – Yikes!







## Rethinking Search

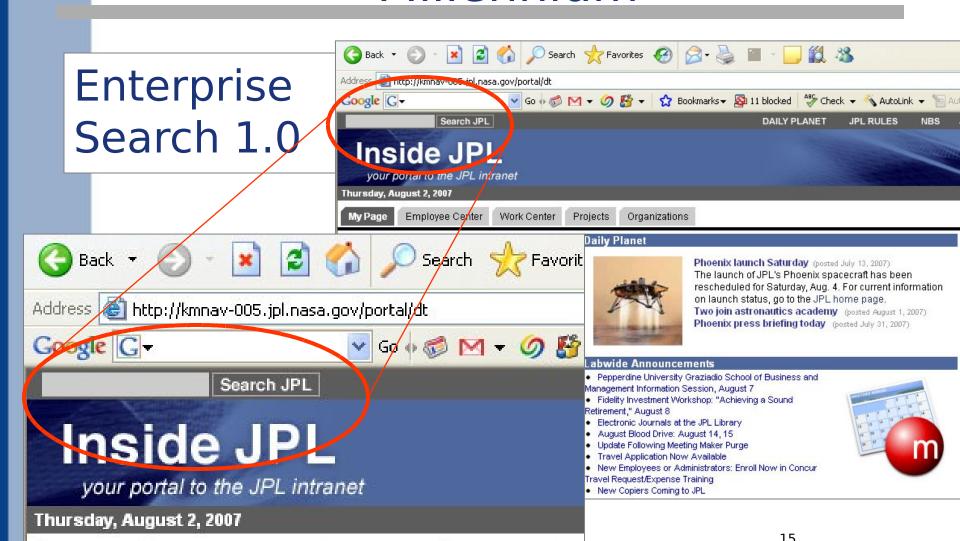
How can we adapt some of the ideas that are out on the Web for Enterprise Search?







#### Search in a Box is So Last Millennium









## Search Evolution: Breaking Out of The Box

- Search is no longer a box in the corner of a web page
  - Web crawling alone doesn't meet the needs of today's businesses
  - Information objects encompass many formats
  - Relationships and context gain new importance
  - Generating metadata must be part of the work process and transparent to the user
  - New technologies present new opportunities







# How Do I Look For Information?

Does Mars Polar Lander = MPL 98? MPL? Mars PL?

- Terms are used inconsistently
- Terms vary from system to system
- Documents and associated Objects are often separated
  - Search is most often time consuming and unsuccessful
  - Design and engineering rationales frequently lost for mission teams of the future

















## Re-imagining Enterprise Search: The High Level Vision

#### The user should be able to:

- Search multiple repositories at once quickly and intuitively
- Find an object without having to know where it is stored
- Use key word queries blended with faceted browse navigation to discover both known and previously unknown data
- Save searches and subscribe to them
- Apply personal tags to content, increasing its findability by other team members
- See relationships between items; increased context improves sense-making
- Use a common, familiar data model based on JPL processes when searching across repositories







#### Use Cases: JPL Engineering Examples

- Mars Science Lab (MSL) Mission Project Manager:
  - "I'd like to see all documents needed to complete my Certification for Flight Readiness and what state they're in, no matter where they are."
- Mission Element Manager:
  - "I need to give a regular report on parts drawings and what their status is so we can track progress and stay on schedule for launch."
- Fabrication Shop Manager:
  - "I need to know the status of drawings so we can be sure to do proper resource planning an be ready to get everything built on time."
- Scientist/Principal Investigator:
  - "I want to see what types of data were returned on earlier missions using a particular type of instrument to help me







## New Strategies for Enterprise Search

- Search across repositories using a common information model
  - Taxonomies, consistent metadata, information services
  - Semantic Web technologies and analytics allow for mappings, filters and logical extensions at the enterprise level
  - Formalize business process to leverage use case analysis
    - Content life cycles, roles, deconstruction of repeating tasks
    - Know your internal customers, find their pain points
  - Patterns will emerge that indicate use cases and workflow possibilities
  - Prep for inferencing and mash ups







## Metadata Development

- 1. Define use cases
  - Determine questions for data analysis
    - What problem are you trying to solve?
    - What relationships need to be expressed?
- Design metadata elements to answer relevant questions
  - Write clear definitions for better understanding
  - Some metadata is mandatory, some is conditional
- 3. If values can be normalized to improve consistency and integrity, consider using controlled vocabulary terms (a taxonomy)







### JPL Taxonomy Effort Goals

- Tightly integrated with JPL Metadata Specification Core
  - Provides vocabulary values for the appropriate metadata fields
- Provides a classification scheme for identifying content
  - A means for tagging content so it can be used and reused in different contexts
- Derived from JPL standard gold sources
  - Vocabulary owners are Subject Matter Experts in their domains
  - Taxonomy maintenance delegated to vocabulary owners
- Integrated with the NASA Taxonomy
  - Ensures interoperability across the Agency







# Getting from Here to There: Mapping Silo'ed Data Structures to the Enterprise Standard

UIMA = Unstructured Information Management Architecture

## **Open Software Architecture and Emerging Standard from IBM**

Platform independent standard for interoperable text and analytics

In Development: UIMA Standards Technical Committee Initiated by OASIS

http://www.oasis-open.org/committees/tc\_home.php?wg\_abbrev=uima

Software Framework Implementation: SDK Available on ABM





#### Do

- Independently developed
- From an increasing # of

#### Analysis Capabilities

- Language, Speaker Identifiers
- Tokenizers
- Classifiers
- Part of Speech Detectors
- Document Structure Detectors
- Parsers, Translators
- Named-Entity Detectors -Taxonomies
- Relationship Detectors

#### Different technologies & interfaces

- specializations
  - Modality
  - Human Language
  - Domain of Interest
  - Source: Style and Format
  - Input/Output Semantics
  - Privacy/Security
  - Precision/Recall Tradeoffs
  - Performance/Precision Tradeoffs...

The right analysis for the job will likely be a <u>best-of-breed</u> <u>combination</u> integrating capabilities across many dimensions.







#### Some UIMA Architecture Terms

#### **Aggregate Analysis Engine**

Contains a collection of component Analysis Engines

#### **Analysis Engine**

A program that analyzes artifacts (e.g. documents) and infers information about them, and which implements the UIMA Analysis Engine interface Specification.

**Annotator** - A software component that implements the UIMA annotator interface. Annotators are implemented to produce and record annotations over regions of an artifact

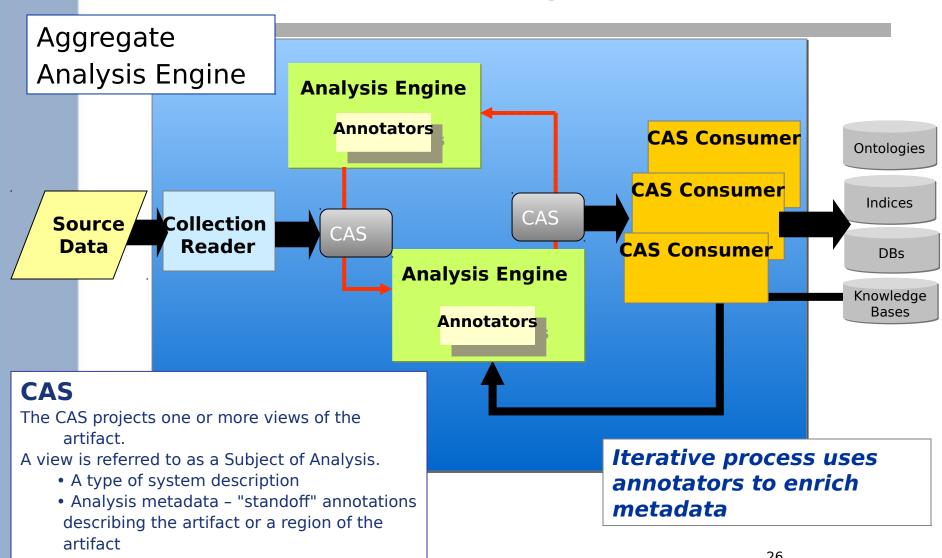
**CAS** - Common Analysis Structure The information object being analyzed







### Generic UIMA Analysis Process

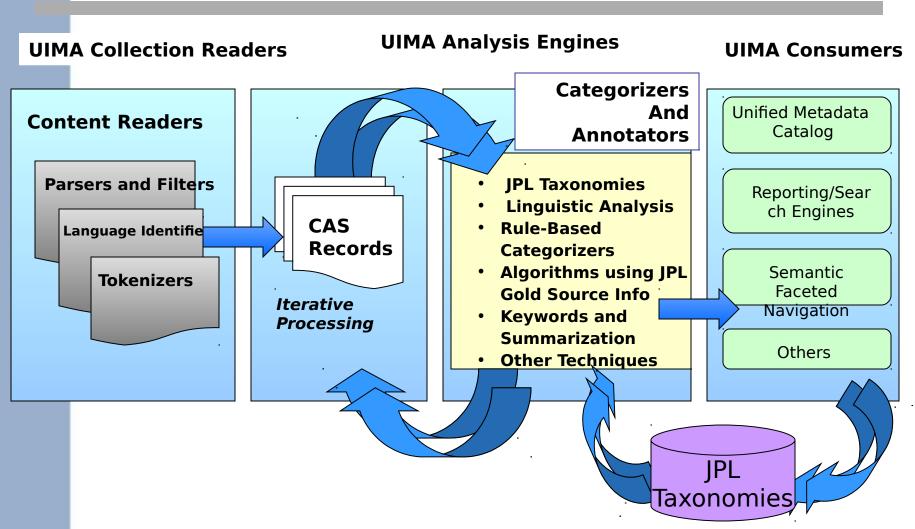


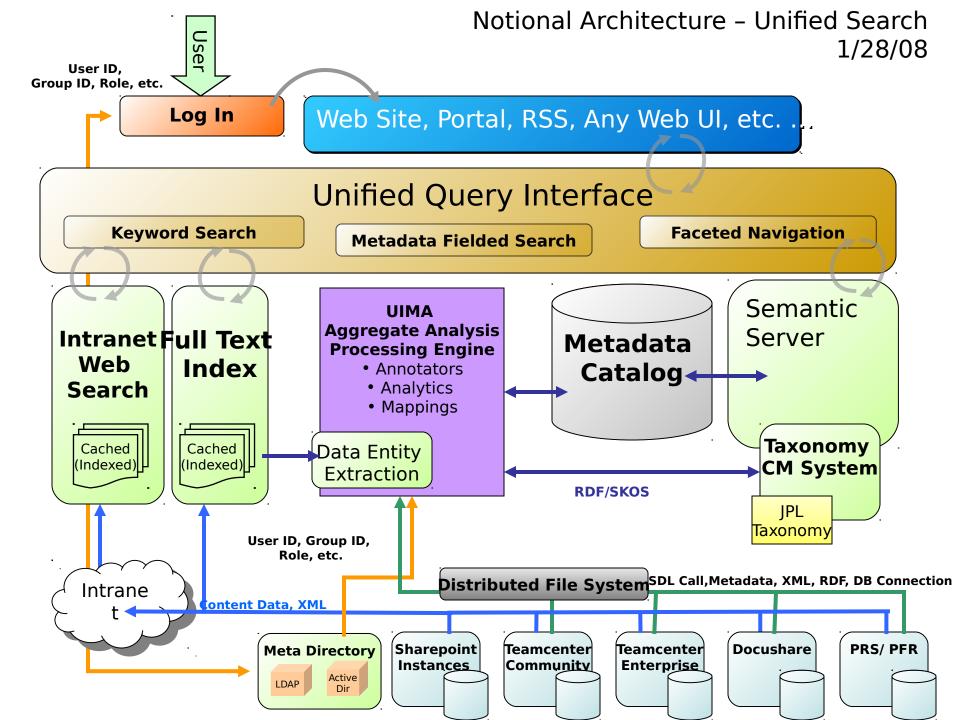






#### UIMA Framework and JPL Unified Search Mappi











# Search Designed by the User

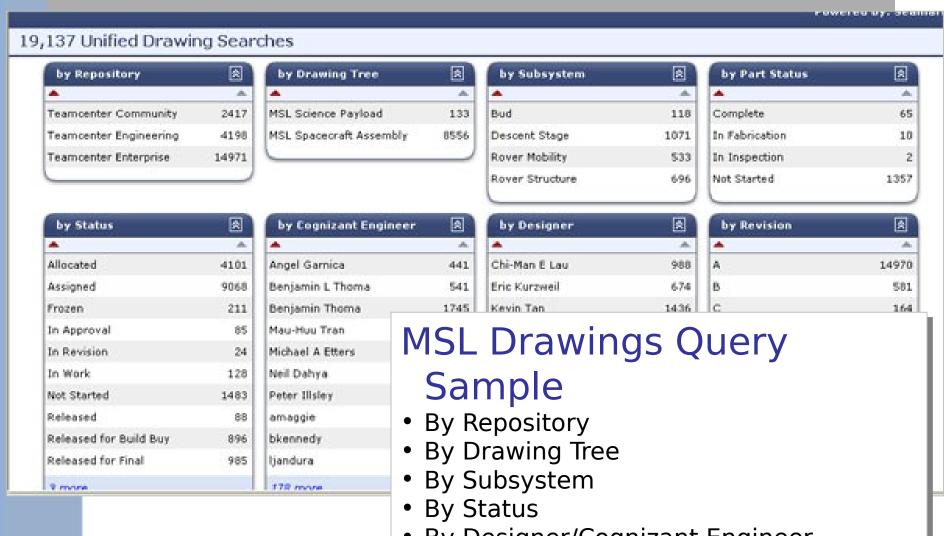
- Pick the search parameters that are of interest
- Save the query
- See the results in text or in pictures
- Subscribe to the search by RSS feed
- Drop the search into your community portlet or wiki page
- Share the query with your team
- Change or add parameters any time







### Unifying Data for Increased Browseability Using Faceted Navigation









#### Search Should be Intuitive



#### Search Strategies

- Browse by Facet (not sure what you're looking for)
- Query by Metadata Field (when one fact or attribute is known)
- Keyword and Full Text (for references in related documents)

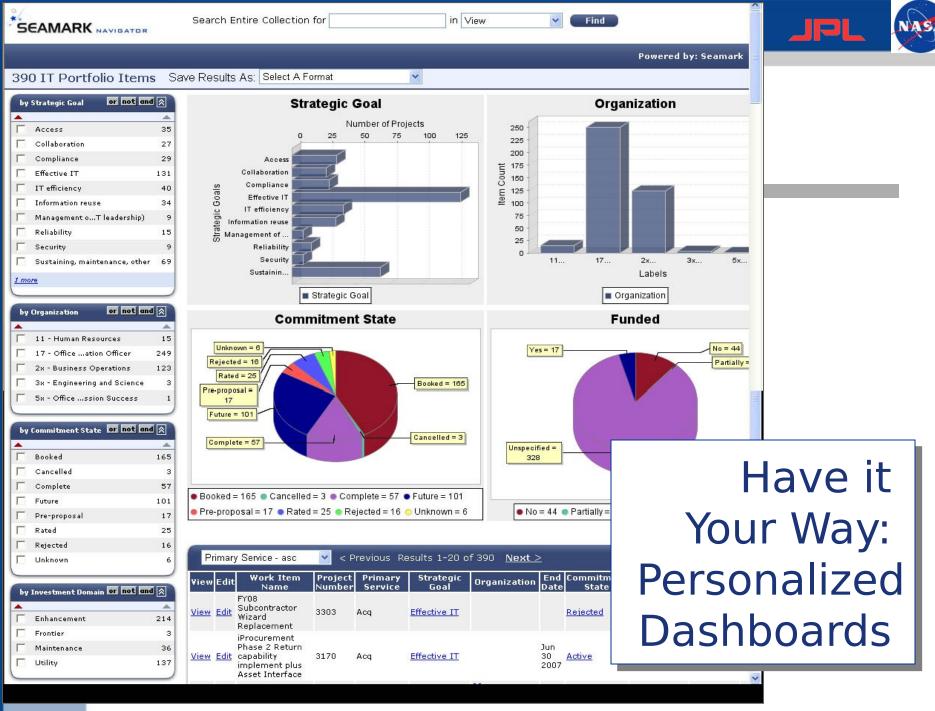






## Managing Results Views

0/	Logged in as <b>administrator</b> <u>Log Ou</u>	t ^
SEAMARK NAVIGATOR	Search by Metadata: All Fields Find	
SCAIVIARK NAVIGATOR	Search by Keyword & Full Text: Find	
	Powered by: Seama	4
	Tomered by Focume	
691 Objects		
	Contest Torse	
Restart	Content Type Most Relevant Pivot to:	
▼ Text of Document contains		
Critical Design Review	□ Tag All All 691 matching items vusing: Add Tag Remove Tag	
Mission / Project: Mars Missions /	Review (691 items in group)	
Program Of fice > Mars Exploration		
Rover	Flight System Implementation Status Review	
Content Type: Review Package	Relevance: 0.5726408362388611	
	Repository: DocuShare	
by Collection	Collection: CD&R - MER Flight System Implementation Status Review 08/10/01	
<u> </u>	Mission / Project: Mars Exploration Rover Mission Phase: Phase C/D, Phase C/D	
Controlled Decords (CD&R) 691	Content Type: Review	
	File Format: application/vnd.ms-powerpoint	
by Organization	Document State: Official	
<u> </u>	System / Subsystem: Flight System, Ground Support Equipment, Integration and Test Equipment,	
101 - SeniorReview Office 1	Payload, Flight Hardware, Flight Software, Guidance, Navigation and Control, Harness, Mechanisms,	
11 - Human Resources 1	Power, Propulsion, Structures, Telecommunications, Thermal, Fault Protection, Entry, Descent and Landing System, Cruise and Entry, Descent and Landing Testbed, Panoramic Camera, Rover	
17 - Officeation Officer 13	Sensitivity: Not Assessed	
2x - Business Operations 3	Date: August 10, 2001	
3x - Engineering and Science 433	Instruments: Rock Abrasion Tool, High Gain Antenna, Low Gain Antenna	
4x - Solar System 7	Competency: Engineering, Science, Avionics, Mission, Technical, Business	
Tax . Solar System	Subject: Engineering, Space Communications, Spacecraft Communications, Command and Tracking,	









## Change Your Viewpoint by Pivoting on Metadata Relationships

<b></b>	_
Assigned	1981
Frozen	18
In Approval	7
In Revision	24
In Signature	10
In Work	127
Not Started	1483
Released	88
Released for Build Buy	22
Released for Final	61
2 more	
THE PERSON NAMED IN COLUMN TWO IS NOT THE PERSON NAMED IN COLUMN TWO IS NA	

by Cognizant Engineer 🙈	
<u> </u>	A
Benjamin D Riggs	200
Benjamin L Thoma	773
Daniel M Coatta	88
John J Quicksall	27
Matthew D Spaulding	27
Michael A Etters	350
Michael J Gradziel	74
Michael W Shafer	30
Peter M Illsley	681

	Cover, Drive Cable	Exit			
	Drawing Number: 10274497, 10274497 Repository: Teamcenter Community, Teamce				
	Drawing Link	Marian Committee of the			
	Subsystem: Rover Mo	obility			
	Part Status: Not Started				
	Project: Mars Science Laboratory, Mars Scien				
	DWG Status: Not Started, Frozen				
	Cognizant Engineer:	Sean J Haggart Q			
	Designer: Albert H Ti	Drawing			
	Category: Wheel Cor	Diaming .			
	Drawing Type: PART	Mobility Assembly,			
	Next-Assembly: Mol	Drawing Number: 102			
	Parent: 10274497-1/	Repository: Teamcen			
	Modified Date: 31-M	Drawing Link, Drawin			
	Part ID: 10274497-D	Subsystem: Rover Mo			
	Dataset Type: Unigra	DWG Tree Location:			
	Owner: Albert H Tran	Mission Phase: Pre A			
	Tag it	Part Status: Not Start			
	Tay it	Project: Mars Science			
	SHIM, WHEEL RES	Create Date: 2005/05			
	Drawing Number: 10	Number of Files: 0			
	Repository: Teamcer	DWG Status: In Work,			
	Drawing Link	Designer: Darrin J Tid			
		Revision: A			
	Subsystem: Rover M Part Status: Not Sta	Category: Top Assem			
	Project Mars Science	Drawing Type: ASSY			

Project: Mars Science

DWG Status: Not St

	Cover, Drive Cable	Exit		
	Drawing Number: 10274497, 10274497			
	Repository: Teamcenter Community, Teamcenter Engineering			
	Drawing Link			
	Subsystem: Rover Mobility Part Status: Not Started Project: Mars Science Laboratory, Mars Science Laboratory DWG Status: Not Started, Frozen			
	Cognizant Engineer: Sean J Haqqart Q			
	Designer: Albert H T	Drawing		
	Category: Wheel Cor	J. Carring .		
	Drawing Type: PART	Mobility Assembly, MSL ROVER MOBILITY DRAWING		
	Next-Assembly: Mol	Drawing Number: 10253201, 10253201		
	Parent: 10274497-1/	Repository: Teamcenter Community, Teamcenter Enterprise		
	Modified Date: 31-M	Drawing Link, Drawing Link		
	Part ID: 10274497-D	Subsystem: Rover Mobility		
	Dataset Type: Unigra	DWG Tree Location: Rover Mobility Installation, Rover Mobility		
	Owner: Albert H Tran	Mission Phase: Pre ATLO		
	Tag it	Part Status: Not Started		
		Project: Mars Science Laboratory, Mars Science Laboratory		
	SHIM, WHEEL RES	Create Date: 2005/05/25-15:43:21:001		
	Drawing Number: 10	Number of Files: 0		
	Repository: Teamcer	DWG Status: In Work, Assigned		
	Drawing Link	Designer: Darrin J Tidwell Q Revision: A		
	Subsystem: Rover M			
	Don't Other tree Hat Cha	Category: Top Assembly		

Next-Assembly: Mobility Assembly Installation Q

INGS

lobility Installation







# Personal Tagging: Creating Folksonomies for Team Use

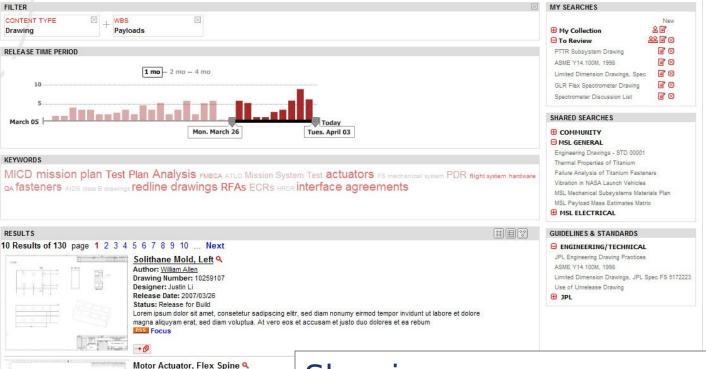






## Next Generation Team Collaboration Platforms







#### Showing:

Author: Richard Frankos Drawing Number: 103104 A,2 Release Date: 2007/03/26 Status: Release for Build

Flex Bender Plug

Author: Horatio Sanzebar Drawing Number: 102414 C,1 Release Date: 2007/03/26

Status: Release for Final

Retaining Plate, Inner

R55 Focus

RSS Focus

Lorem ipsum dolor sit amet, consetetur sadipscing elit magna aliquyam erat, sed diam voluptua. At vero eos

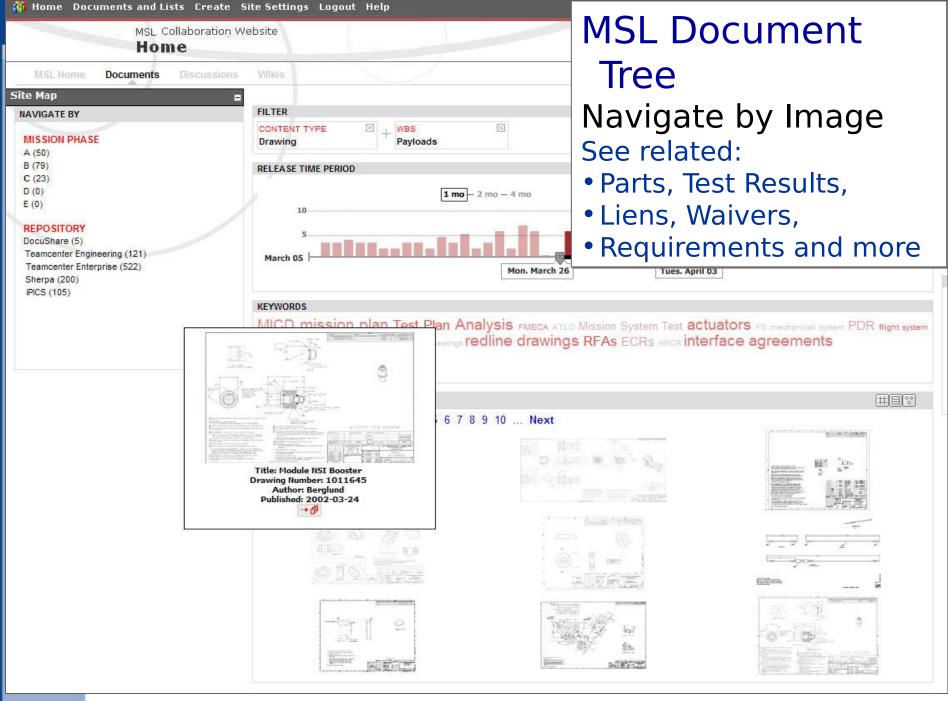
Lorem ipsum dolor sit amet, consetetur sadipscing elit

magna aliguvam erat, sed diam voluptua. At vero eos

sit amet, consetetur sadipscing elitr, sed diam nonumy

erat, sed diam voluptua. At vero eos et accusam et ju

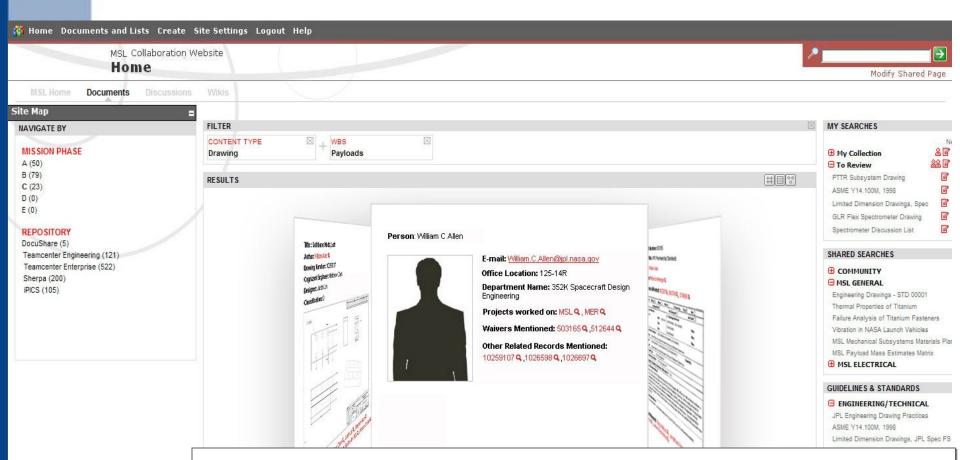
- Saved searches
- Team searches
- Subscribed searches
  - Search by sliding parameter sca
- Faceted navigation for browse











#### Look At Multiple Objects in Context

- Includes data on people, projects as well as engineering products
- Modern visualization and improved interface







## Summing Up

- Web 2.0 presents many new opportunities to redesign search
- What's new is socialization and increased interactions between content and people
- Context gains importance
  - Relationships help establish context
- We used to call it search; now we say:
  - Enabling social networking across teams
  - Information discovery without having to know an exact key word
  - Personal reporting quickly designed and easily accessible
  - Flexible graphics and sexy interfaces that resonate with the iPod generation







## Questions and Discussion

# Thanks for your time and interest!

Jayne.E.Dutra@jpl.nasa.gov







## Back Up Information







## Enterprise Data Integration Using UIMA

### **Open Software Architecture and Emerging Standard from IBM**

Platform independent standard for interoperable text and analytics

Under Development: UIMA Standards Technical Committee Initiated under OASIS

http://www.oasis-open.org/committees/tc\_home.php?wg\_abbrev=uima

#### **Software Framework Implementation**

SDK Available on IBM Alphaworks

http://www.alphaworks.ibm.com/tech/uima

Tools, Utilities, Runtime, Extensive Documentation

Creation, Integration, Discovery, Deployment of analytics Java, C++, Perl, Python (others possible)

Supports co-located and service-oriented deployments (eg., SOAP)

x-Language High-Performances APIs to common<sub>4</sub>data